Immunogenetics (2005) 57: 590–600
DOI 10.1007/s00251-005-0016-y

ORIGINAL PAPER

Takako Maruoka · Hideyuki Tanabe · Marin Chiba · Masanori Kasahara

Chicken *CD1* genes are located in the MHC: *CD1* and endothelial protein C receptor genes constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals

Received: 19 May 2005 / Accepted: 20 June 2005 / Published online: 26 August 2005

Abstract Mammals have several major histocompatibility complex (MHC) class-I-like genes. Although some of them are assumed to have originated before the emergence of mammals, the origin of class-I-like genes is poorly understood. We analyzed here the recently released chicken draft genome sequence and identified two families of class-I-like genes: *CD1* and *PROCR* (the gene for the endothelial protein C receptor). Chickens have two *CD1* genes, designated *CD1.1* and *CD1.2*, located in tandem ~840 bp apart from each other. Chicken *CD1.1* and *CD1.2* are neither group 1- nor group 2-like, indicating that the two groups of *CD1* emerged in a mammalian lineage. Although the database provides no information as to their chromosomal localization, we found that chicken *CD1* genes are located adjacent to the previously characterized MHC *B* system contig on chromosome 16. We confirmed the linkage of *CD1* to the *B* system by dual-color fluorescence in situ hybridization. Chickens have a single copy of *PROCR*. Among known class-I-like genes, *PROCR* is most closely related to *CD1*, indicating that *CD1* and *PROCR* constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals.

Sequence data reported in this paper have been submitted to the DDBJ/EMBL/GenBank databases under accession numbers AB204797–AB204802, AB205394, and AB205528

Electronic supplementary material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00251-005-00016-y and accesible for authorized users.

T. Maruoka · M. Kasahara (✉)
Department of Pathology, Division of Pathophysiological Science, Hokkaido University Graduate School of Medicine, North-15 West-7,
Sapporo, 060-8638, Japan
e-mail: mkasaha@med.hokudai.ac.jp
Tel.: +81-11-7065048
Fax: +81-11-7067825

T. Maruoka · H. Tanabe · M. Chiba · M. Kasahara
Department of Biosystems Science,
School of Advanced Sciences,
The Graduate University for Advanced Studies (Sokendai),
Hayama, 240-0193, Japan

Keywords MHC · CD1 · Endothelial protein C receptor · Chicken

Introduction

Classical class I genes of the major histocompatibility complex (MHC) are highly polymorphic and expressed almost ubiquitously (Klein 1986). They encode the heavy chain of class I molecules that present antigenic peptides to $CD8^+$ T cells. In addition to classical class I genes, mammalian genomes contain several class-I-like genes collectively known as nonclassical class I genes. Although the majority of nonclassical class I genes located in the MHC show reasonably high sequence similarity to classical class I genes with the exception of *MICA/B* (Bahram et al. 1994) and *HFE* (the gene for hemochromatosis) (Feder et al. 1996), most of those encoded outside the MHC are quite divergent from classical class I genes (Radosavljevic and Bahram 2003). Among such divergent class-I-like genes are *CD1* (Calabi and Milstein 1986), *AZGP1* (the gene for zinc-α2-glycoprotein) (Araki et al. 1988), *FCGRT* (the gene for the heavy chain of the neonatal IgG Fc receptor) (Simister and Mostov 1989), *PROCR* (the gene for the endothelial protein C receptor) (Fukudome and Esmon 1994), *MR1* (Hashimoto et al. 1995), *RAET/ULBP* (the genes for retinoic acid early transcripts 1) (Zou et al. 1996), and *Mill* (MHC class-I-like genes located near the leukocyte receptor complex) (Kasahara et al. 2002; Watanabe et al. 2004). Accumulated evidence indicates that class-I-like genes have diverse functions ranging from specialized antigen presentation (Bahram 2000; Treiner et al. 2003; Vincent et al. 2003) to the transport of IgG (Simister and Rees 1985) and lipid mobilization and catabolism (Todorov et al. 1998).

Although it is generally assumed that some class-I-like genes, in particular those encoded outside the MHC, originated prior to the emergence of mammals, their origin is poorly understood. Recently, the draft genome sequence of the chicken, with 6.6-fold coverage, was deposited in the publicly available databases (Hillier et al. 2004). Among nonmammalian vertebrates for which draft genome se-

quences are available, chickens are phylogenetically most closely related to mammals, hence serving as a good starting point for searching the origin of class-I-like genes. In the present study, we searched the chicken draft sequence for class-I-like genes and found that chickens have *CD1* and *PROCR* families of class-I-like genes. We also found that, unlike mammalian *CD1*, chicken *CD1* genes are located in the MHC.

Materials and methods

Isolation of CD1.1 and CD1.2 cDNA by RACE

We performed a computer search of the Ensembl (http://www.ensembl.org/) and National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) databases. Based on the information obtained from the databases, we designed primers for rapid amplification of cDNA ends (RACE) experiments. Briefly, total cytoplasmic RNA isolated from the spleen of a chicken obtained from a local dealer was converted to cDNA and subjected to RACE using the GeneRacer kit following the instructions of the manufacturer (Invitrogen, Carlsbad, CA, USA). This kit allows selective identification of full-length mRNAs by removing the 7-methylguanine cap and replacing it with specific oligo (Maruyama and Sugano 1994). The gene-specific primers for 3′-RACE were 5′-CGCGTTTTCCAG ATCCACACAGG-3′ and 5′-CCAACGGAACGAGATGG AGCTTTG-3′ for CD1.1 and 5′-AGCTCCAGTCCAACG GGACCATC-3′ and 5′-CCAACGGGACCATCAGGACC TTC-3′ for CD1.2. The gene-specific primers for 5′-RACE were 5′-TCTCTCCAGATCTGTCCTCCCTTCC-3′ and 5′-GGAGGAGGAGAAGACGTGCTCCA-3′ for CD1.1 and 5′-TCTGCTTTCCCAGCCTGGATGAA-3′ and 5′-TGTC CACACAGGTGTCGTTGAGGA-3′ for CD1.2. Polymerase chain reaction (PCR) products obtained by 3′- and 5′-RACE were cloned into the pGEM-T Easy vector (Promega, Madison, WI, USA), and their sequences were determined using an automated sequencer. Potential PCR errors were eliminated by setting up PCR reactions in triplicate and sequencing multiple clones for each reaction.

Isolation of PROCR cDNA by RACE

A computer search of the Ensembl and NCBI databases resulted in the identification of high-throughput genomic and expressed sequence tag (EST) sequences predicted to encode chicken PROCR. We designed primers for RACE experiments based on the EST sequences, and total RNA isolated from the thymus was converted to cDNA using the GeneRacer kit. Gene-specific primers for 3′-RACE were 5′-ATGCTCAGGCTGCTGCTGCT-3′ and 5′-TCAGCGG CATAGTGCAGGTC-3′, and those for 5′-RACE were 5′-AGGATGTCCACGCAGGTGGT-3′ and 5′-GGGACGTG GAAGGTGAGGAA-3′. Cloning and sequencing were performed as described above.

Determination of gene structures and analysis of surrounding genomic regions

In principle, exon–intron structures of chicken *CD1* and *PROCR* genes were deduced by comparing cDNA sequences determined above with the genomic sequences deposited in the NCBI (build 1.1) and Ensembl (29.1e) databases. For those regions where the draft genome sequence was absent or judged to contain assembly errors, we obtained the sequences using PCR products amplified from the chicken genomic DNA. The interval between *CD1.1* and *LOC427729* (the gene coding for a putative G-protein-coupled receptor) was amplified by PCR (initial denaturation of 4 min at 91°C followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 58°C for 1 min and extension at 72°C for 1.5 min, and a final extension at 72°C for 5 min) using the primers 5′-ACCTCATCCCTTCTCCTAC-3′ and 5′-CCGTAGGACT GTCACCCACT-3′. The interval between *LOC427729* and *TN-Y* (the gene for tenascin-Y, the chicken counterpart of tenascin-X, which in mammals is encoded in the class III region) (Hagios et al. 1996) was amplified by long-range, nested PCR using two sets of primers: 5′-ATGACACAGCTCCACAGCAC-3′ and 5′-GGGACAC ACACAGGACACAA-3′ (first round) and 5′-CAGAGA CAGTCGCATTGGAA-3′ and 5′-CAGGCTCTGAGGGC TCAATA-3′ (second round). Cycling conditions were initial denaturation of 1 min at 94°C followed by 30 cycles of denaturation at 98°C for 10 s and annealing and extension at 68°C for 15 min (10 min for nested PCR) and a final extension at 72°C for 10 min. The interval between *TN-Y* and *histone H3* was amplified by long-range PCR using three sets of primers: 5′-GAG TACTCCATCCCCTTCACAGAGATGA3′ and 5′-CTTT ATCACCGCTGTTTTCCCGATTT-3′ (first round), 5′-GG GTCGCAGTGGAGCCTTTA-3′ and 5′-CCCCAAAAAT GGGACTTACC-3′ (second round), and 5′-GGAGCCTT TATGGGGTCAAT-3′ and 5′-CCCCAAAAATGGGACT TACC-3′ (third round). Cycling conditions were initial denaturation of 1 min at 94°C followed by 30 cycles of denaturation at 98°C for 10 s and annealing/extension at 63°C for 15 min and a final extension at 72°C for 10 min.

Dual-color FISH

Chicken tapasin cDNA was isolated by nested reverse transcription–PCR (RT-PCR) from liver cDNA using the forward primer (5′-GCTCTGCTGGTCCCAATTTA-3′) and two sets of reverse primers (5′-GGTCACTGCGATTTCTTGGT-3′ and 5′-TAGGGTAGAGCCAACGGATG-3′). These primers were designed on the basis of the published cDNA sequence (Frangoulis et al. 1999) and the draft genome sequence. Cycling conditions were initial denaturation of 4 min at 91°C followed by 30 (first round) or 25 (second round with nested primers) cycles of denaturation at 94°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 1–1.5 min and a final extension at 72°C for 5 min.

Identity of the tapasin cDNA was confirmed by partial DNA sequencing.

The fragment lengths of CD1 and tapasin cDNA probes were modified by DNase I digestion to the size ranging from 300 to 600 bp. The digested CD1 and tapasin cDNA fragments were labeled with Alexa Flour 532 and Oregon Green 488 using ULYSIS Nucleic Acids Labeling Kits (U21659 and U21651, respectively) according to the manufacturer's instructions (Molecular Probes, Eugene, OR, USA). Two micrograms of labeled DNA was mixed with 10 μg of chicken Cot-1 DNA and dissolved in 8 μl of hybridization solution (50% formamide and 10% dextran sulfate in 2×SSC). The hybridization mixtures were applied onto the metaphase spreads prepared from the chicken cell line DT40 (derived from JCRB9130). Fluorescence in situ hybridization (FISH) procedures were followed by the standard protocols described elsewhere (Tanabe et al. 2005). The slides were counterstained with 4′,6-diamidino-2-phenylindole (1 μg/ml) and mounted with antifade medium Vectashield (Vector Laboratories, Burlingame, CA, USA). Monitoring of the fluorescent twin spot signals on the metaphases was performed under an epifluorescence microscopy (Leica DM5000B) equipped with single-band-pass filters. FISH images were captured with a Leica DFC 350 FX camera and analyzed by the Leica CW4000 software program.

Sequence analysis

Signal peptides were predicted using the SignalIP version 2.0.b2 server (http://www.cbs.dtu.dk/services/SignalP-2.0/). Transmembrane segments (TMs) were predicted using the SOSUI program (http://www.sosui.proteome.bio.tuat.ac.jp/sosui_submit.html). Kyte and Doolittle hydropathy plots (Kyte and Doolittle 1982) with a window size of nine residues were generated using the ProtScale program (http://www.us.expasy.org/tools/protscale.html). Amino acids sequences were aligned with the Clustal W program (Thompson et al. 1994). The alignment was then adjusted by eye to maximize sequence similarity. For phylogenetic analysis, the distance matrix was obtained by calculating p distances for all pairs of sequences. Sites containing gaps were excluded from the analysis using the pairwise-deletion option. Neighbor-joining trees were constructed using the MEGA software version 3.0 (Kumar et al. 2004). The reliability of branching patterns was assessed by bootstrap analysis (5,000 replications).

Fig. 1 Amino acid sequences of chicken CD1.1 and CD1.2 deduced by cDNA cloning. The signs "–" and "/" indicate identity with the top sequence and absence of residues, respectively. Predicted signal peptides and TM are underlined. *Triangles* indicate exon–intron boundaries. *Diagonals* in the α2 and α3 domains indicate cysteine residues presumed to form the disulfide bridge. Potential *N*-linked glycosylation sites (two in CD1.1 and four in CD1.2) are indicated by *thick underlines*. Potential sorting motifs in the CYT are *boxed*. The CYT sequences of human CD1B and CD1C are also shown to highlight the motifs. Chicken CD1 cDNA sequences have been submitted to the DDBJ/EMBL/GenBank databases under accession numbers AB204802 (CD1.1) and AB204801 (CD1.2). For accession numbers of human CD1 molecules, see the legend of Fig. 2

Results

Isolation of chicken CD1.1 and CD1.2 cDNA

We performed tBLASTn searches of the Ensembl and NCBI chicken draft genome databases using the human CD1 sequences (P15813 and AAA51931.1) as queries. Two high-throughput genomic sequences (XM_423513 and XM_423514), presumed to contain *CD1* genes, designated *CD1.1* and *CD1.2*, were identified. To facilitate cDNA cloning, we first deduced the α1–α3 domain sequences of chicken CD1.1 and CD1.2 from the draft genome sequence and searched the EST database using these sequences as queries. This led to the identification of EST encoding CD1.1 (CD2181331, CB0170461, BG710491, BG7131691, CB0173851, and CK6131741) and CD1.2 (AJ396603 and AJ442905). We designed gene-specific primers based on this information and isolated full-length cDNA by RACE. Figure 1 shows the amino acid alignment of chicken CD1 and human CD1A/CD1D. Deduced chicken CD1.1 and CD1.2 molecules are made up of 348 and 374 amino acids, respectively. Both molecules were predicted to have an N-terminal signal peptide consisting of 17 amino acids and a single TM. Hence, mature forms of chicken CD1.1 and CD1.2 are predicted to have 331 and 357 amino acids, respectively. In the α2 domain of CD1.1, one of the cysteine residues involved in disulfide bridge formation was changed to phenylalanine. Interestingly, this cysteine residue is changed to tryptophan in the α2 domain of mouse CD1d2. Comparison of chicken CD1.1 and CD1.2 showed that their α3 domains are almost identical (98% amino acid identity), whereas the remaining extracellular domains are poorly conserved (25 and 22% amino acid sequence identity for the α1 and α2 domains). Many mammalian CD1 molecules contain tyrosine-based endosomal sorting motifs in their cytoplasmic tails (CYT) (Sugita et al. 2004). The CYT of chicken CD1.2 contains a potential tyrosine-based motif. On the other hand, chicken CD1.1 does not contain such motifs in its CYT, but its C-terminus ends with leucine–isoleucine residues that may function as a modified dileucine motif (Sandoval and Bakke 1994). We compared hydropathy plot profiles of the α1 and α2 domains of chicken CD1.1 and CD1.2 with those of chicken and human classical class I molecules and human CD1 molecules (see the electronic supplementary figure 6). This analysis revealed that, similar to human CD1, the α1 and α2 domains of chicken CD1.1 and CD1.2 were more hydrophobic than those of classical class I molecules.

BLAST searches of the NCBI database identified CD1 of various mammalian species as molecules most similar to chicken CD1.1/CD1.2 (30% amino acid identity). Phylogenetic analysis (Fig. 2) confirmed that chicken CD1.1 and CD1.2 qualify as members of the CD1 family. Notably, chicken CD1.1 and CD1.2 were more closely related to each other than they were to any of the mammalian CD1 molecules. Thus, the duplication event that gave rise to chicken *CD1.1* and *CD1.2* appears to have taken place independently from the one that created the two groups of mammalian *CD1* genes.

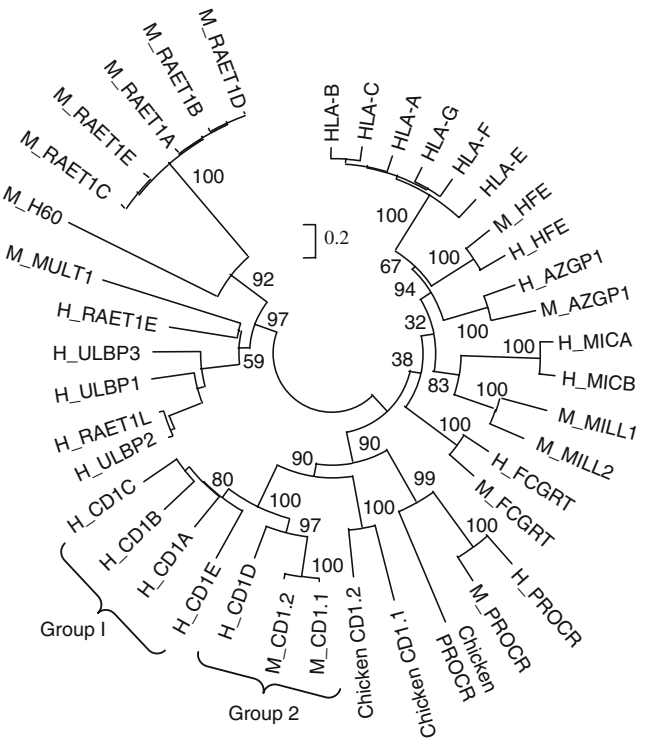

Fig. 2 Neighbor-joining tree of chicken and representative mammalian class I and class-I-like molecules. Accession numbers of the sequences were as follows: HLA-A, HLHUA2; HLA-B, AF189017; HLA-C, NP_002108; HLA-E, NP_005507; HLA-F, NP_061823; HLA-G, A39953; H_HFE, P70387; M_HFE, Q30201; H_AZGP1, BC005306; M_AZGP1, D21059; H_MICA, A55739; H_MICB, U65416; M_MILL1, AB086265; M_MILL2, AB086267; H_FCGRT, BC008734; M_FCGRT, L17022; H_PROCR, BC014451; M_PROCR, A55945; M_CD1.1, M63695; M_CD1.2, M63697; H_CD1A, AAA51931; H_CD1B, BC069481; H_CD1C, P29017; H_CD1D, B C027926; H_CD1E, X14975; H_RAET1E, AAL76417; H_RAET1L, AAK91503; H_ULBP1, NP_079494; H_ULBP2, NP_079493; H_ULBP3, NP_078794; M_RAET1A, 2206404A; M_RAET1B, NP_033043; M_RAET1C, NP_033044; M_RAET1D, NP_064414; and M_RAET1E, AAL11004. H and M stand for human and mouse, respectively. Bootstrap values are shown only at major nodes

Structures of the chicken *CD1* genes

Supercontig NW_090165.1, made up of contigs 3696.1 and 3696.2, contained the entire exons of the *CD1.2* gene (Fig. 3a). We deduced the exon–intron organization of this gene by comparing our cDNA sequence with the sequence of the supercontig. Exon 1, intron 1, exon 2, and part of intron 2 of *CD1.2* were in contig 3696.1, and part of intron 2 and the rest of the gene were in contig 3696.2. Supercontig NW_090165.1 did not contain the whole intron 2 sequence because contigs 3696.1 and 3696.2 were separated by a gap. We filled this gap by sequencing genomic PCR products (accession number AB204797). This showed that intron 2 was made up of 885 bp. Contig

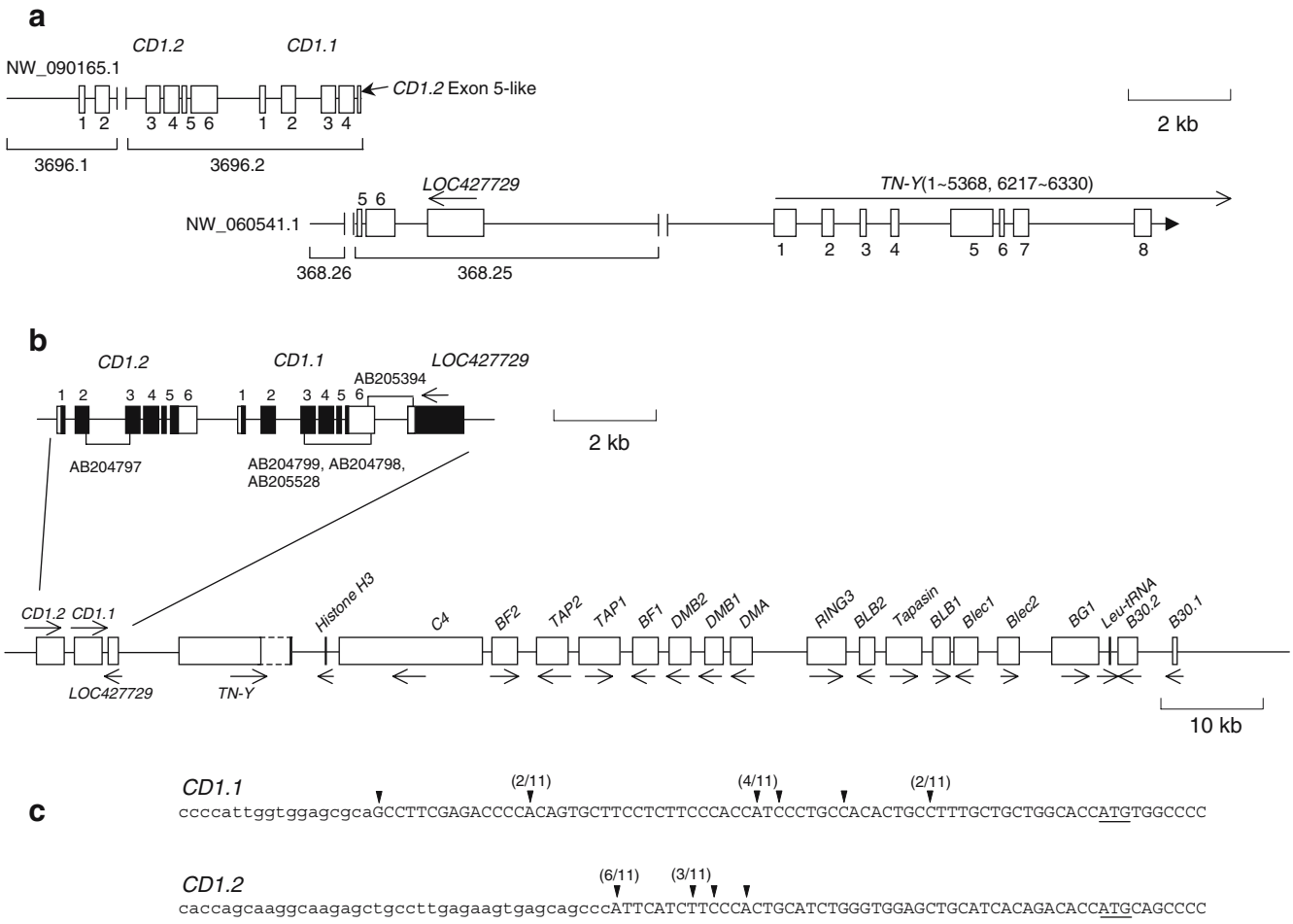

Fig. 3 **a** Schematic maps of relevant contigs deposited in the public databases. The draft sequence NW_090165.1 is made up of two contigs, with a gap (indicated by *vertical lines*) in intron 2 of *CD1.2*. The draft sequence NW_060541.1 is made up of 26 contigs, of which only two contigs, 368.26 and 368.25, are shown in this figure. **b** Organization of the extended chicken MHC. Exon–intron organization is shown for *CD1* and *LOC427729*. *Numbered boxes* indicate exons. Although the full-length cDNA sequence is not available, *LOC427729* appears to have only one exon. Coding regions are shown as *filled boxes*. *Empty boxes* indicate 5′- and 3′-UTR. Genes were named according to the recent recommendation (Miller et al. 2004). Regions sequenced in this work are indicated by *brackets* along with database accession numbers. **c** Transcription initiation sites of chicken *CD1* genes. *Arrowheads* on top of the sequences indicate the start positions of the 5′-RACE clones. For each gene, 11 clones were sequenced. The number of clones that started at each position is indicated in *parentheses* when there are multiple clones. Exon 1 is shown in capital letters

3696.2 contained exons 1–4 of the *CD1.1* gene, but lacked exons 5–6. We therefore obtained the sequence of this region including exon 3 to intron 4 (accession numbers AB204798, AB204799, and AB205528) and deduced the structure of the *CD1.1* gene.

The two *CD1* genes are located in tandem, only ~840 bp apart from each other (Fig. 3b). They have identical exon–intron organization and are compact in size (~2.8 kb for *CD1.1* and ~2.9 kb for *CD1.2*). Exon 1 codes for the 5′-untranslated region (UTR), the signal peptide, and the first four residues of the α1 domain. The α1, α2, and α3 domains are encoded by exons 2, 3, and 4, respectively. Exon 5 encodes the connecting peptide (CP), TM, and part of CYT, and exon 6 encodes the CYT and 3′-UTR. In both *CD1.1* and *CD1.2*, all exon–intron boundaries conformed to the canonical guanine–thymine/adenine–guanine (GT/AG) rule without exception. Like mammalian *CD1* genes, RNA splicing took place between the first and second base of the junctional codons in all boundaries, with the exception of intron 5 where splicing took place between the second and third base. Analysis of EST sequences indicated that *CD1.2* has a splicing variant where intron 3 (made up of 87 bp) is not spliced out, resulting in the insertion of 29 amino acids between the α2 and α3 domains (clone 27b4r1 derived from the bursa of Fabricius; accession number AJ396603). However, we were unable to detect this splicing variant when we performed PCR using spleen cDNA as a template (data not shown).

We mapped the transcription initiation sites of the chicken *CD1.1* and *CD1.2* genes by sequencing 11 clones each of 5′-RACE products (Fig. 3c). In *CD1.2*, six clones started at the adenine residue 45 bp upstream from the translation initiation site. In *CD1.1*, transcription initiation sites were more variable, suggesting that this gene lacks

specific major transcription initiation sites. The 5′-flanking region of neither *CD1.1* nor *CD1.2* contained a canonical TATA or CCAAT box.

Chicken *CD1* genes map to the MHC on chromosome 16

The draft genome sequence provided no information as to the chromosomal localization of *CD1* genes, because supercontig NW_090165.1 containing *CD1.2* and part of *CD1.1* was not localized to any specific chromosome and was not linked to any other contigs with known chromosomal localization (Fig. 3a). Surprisingly, however, a BLAST search of the Ensembl genome assembly using the CD1.1 cDNA nucleotide sequence (AB204802) as a query revealed that the 3′-UTR sequence of CD1.1 cDNA overlaps with that of contig 368.25 located at the end of supercontig NW_060541.1, encoding the chicken MHC (the *B* system) on chromosome 16 (Bloom and Bacon 1985; Fillon et al. 1998; Kaufman et al. 1999a,b). Indeed, detailed analysis showed that contig 368.25 contains the sequence with 98% identity with that of exon 5, intron 5, and exon 6 of *CD1.1*. Thus, *CD1.1* is located across the two supercontigs NW_090165.1 and NW_060541.1.

Immediately downstream from the *CD1.1* exon 4 sequence, contig 3696.2 contains a sequence (~220 bp) that is ~80% identical to that of intron 4, exon 5, and intron 5 of *CD1.2* at the nucleotide level; however, this sequence was not identified by our genomic PCR experiments (accession numbers AB204798 and AB204799). Because the exon-5-like sequence in contig 3696.2 contains a frameshift mutation and lacks consensus splice acceptor/donor sequences, it is likely derived from a *CD1* pseudogene or gene fragment. We presume that this ~220-bp sequence stretch was misassigned to contig 3696.2. Our genomic PCR experiments also ruled out the possibility that the sequence corresponding to contig 368.26 (located at the very end of supercontig NW_060541.1; Fig. 3a) is located between exons 4 and 5 of the *CD1.1* gene.

Comparison of the draft sequence (NW_060541.1) and the sequence of the chicken MHC *B* system reported by Kaufman et al. (1999b) indicates that the former contains several gaps, but is colinear with the latter in overlapping regions. Supercontig NW_060541.1 extends downstream of the *histone H3* pseudogene located at the end of the sequence of Kaufman et al. (1999b) and contains *TN-Y* and *LOC427729* (Fig. 3b). Because supercontig NW_060541.1 contains gaps between *histone H3* and *TN-Y* and between *TN-Y* and *LOC427729*, we performed genomic PCR to estimate the gap size and to confirm the physical linkage of these genes. We found that *histone H3* and *TN-Y* are ~3.5 kb apart and that the distance between *TN-Y* and *LOC427729* is ~5.7 kb. Similarly, we confirmed by genomic PCR and sequencing that *CD1.1* and *LOC427729* are ~600 bp apart from each other (accession number AB205394). Because *histone H3* has been localized to the MHC (Kaufman et al. 1999b), these results provide convincing evidence that, unlike human and mouse *CD1* genes, chicken *CD1* genes are located adjacent to the class I (*BF*) and II*B* (*BLB*) genes and thus encoded within the MHC *B* system. Orientation of the cosmid cluster (AL023516.1) relative to the previously characterized class II*A* gene (*BLA*) (Kaufman et al. 1995; Salomonsen et al. 2003) indicates that the *CD1* genes map between *BF* and *BLA*.

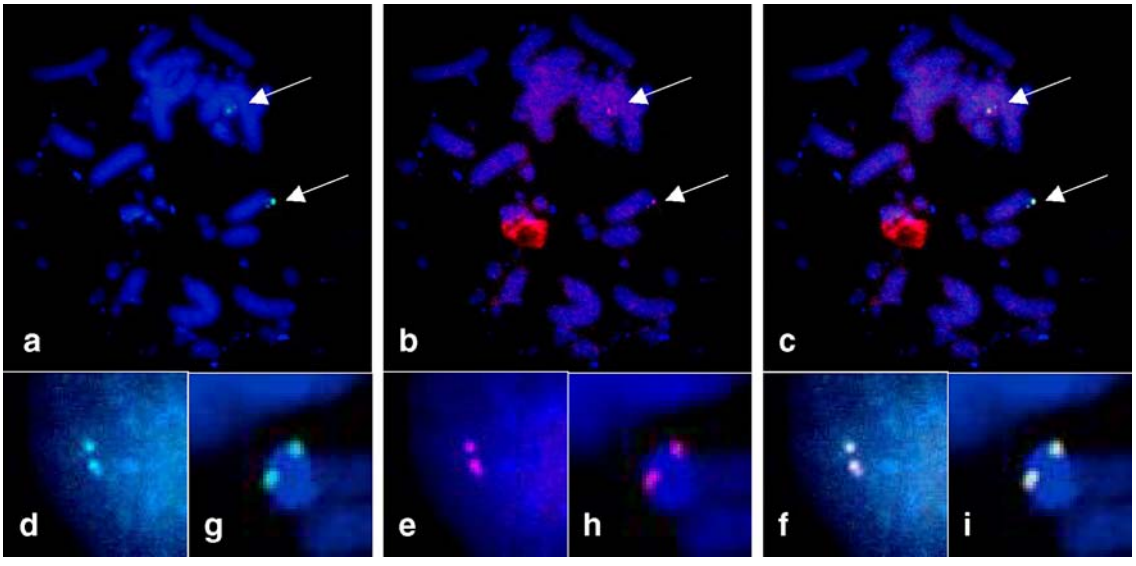

Fig. 4 Chromosomal localization of chicken *CD1* by dual-color FISH. *Tapasin* and *CD1* genes, which are represented by green and red signals, respectively, were cytogenetically colocalized on the distal part of the long arm of a medium-sized chromosome. DNA counterstaining with 4′,6-diamidino-2-phenylindole is shown in blue. Panels **a**, **d**, and **g** show green signals of the *tapasin* gene. Panels **b**, **e**, and **h** show red signals of the *CD1* genes. Panels **c**, **f**, and **i** indicate colocalization of green and red signals. Panels **a**, **b**, and **c** show metaphase images with twin spots indicated by *arrows*. Panels **d**, **e**, and **f** show images of part of interphase nuclei with twin spots. Panels **g**, **h**, and **i** show a single chromosome of metaphase spreads with twin spots on the distal part of its long arm

To independently test the physical linkage of *CD1* to the MHC, we performed dual-color FISH using chicken tapasin cDNA and a mixture of chicken CD1.1 and CD1.2 cDNA as probes (Fig. 4). The red signal detecting *CD1.1* and *CD1.2* was colocalized to the green signal detecting the *tapasin* gene located in the MHC (Frangoulis et al. 1999; Kaufman et al. 1999b), consistent with the conclusion that *CD1.1* and *CD1.2* are encoded in the *B* system.

Isolation of chicken PROCR cDNA and the organization of the PROCR gene

Another family of MHC class-I-like genes identified through BLAST searches of the chicken draft genome was *PROCR* coding for the endothelial protein C receptor. A computer search of the EST database at NCBI identified three *PROCR* sequences (BU303961, BU302284, and BU393787). On the basis of the information obtained from the draft genome and EST sequences, we designed gene-specific primers and isolated full-length PROCR cDNA by RACE. The deduced chicken PROCR molecule is made up of 223 amino acids, of which the N-terminal 16 residues were predicted to constitute the signal peptide (Fig. 5a). Thus, the protein moiety of the mature form of PROCR has 207 amino acids, with a calculated molecular mass of 23,018.39. Chicken PROCR was predicted to have a single TM and seven *N*-linked glycosylation sites. Amino acid sequences of chicken PROCR were 40, 38, 35, and 34% identical to those of human, bovine, mouse, and rat PROCR, respectively.

We determined the exon–intron organization of chicken *PROCR* by comparing the cDNA sequence with the draft genome sequence deposited in the database (XM_422681). The exonic sequence predicted from the database was completely identical to our cDNA sequence. Chicken

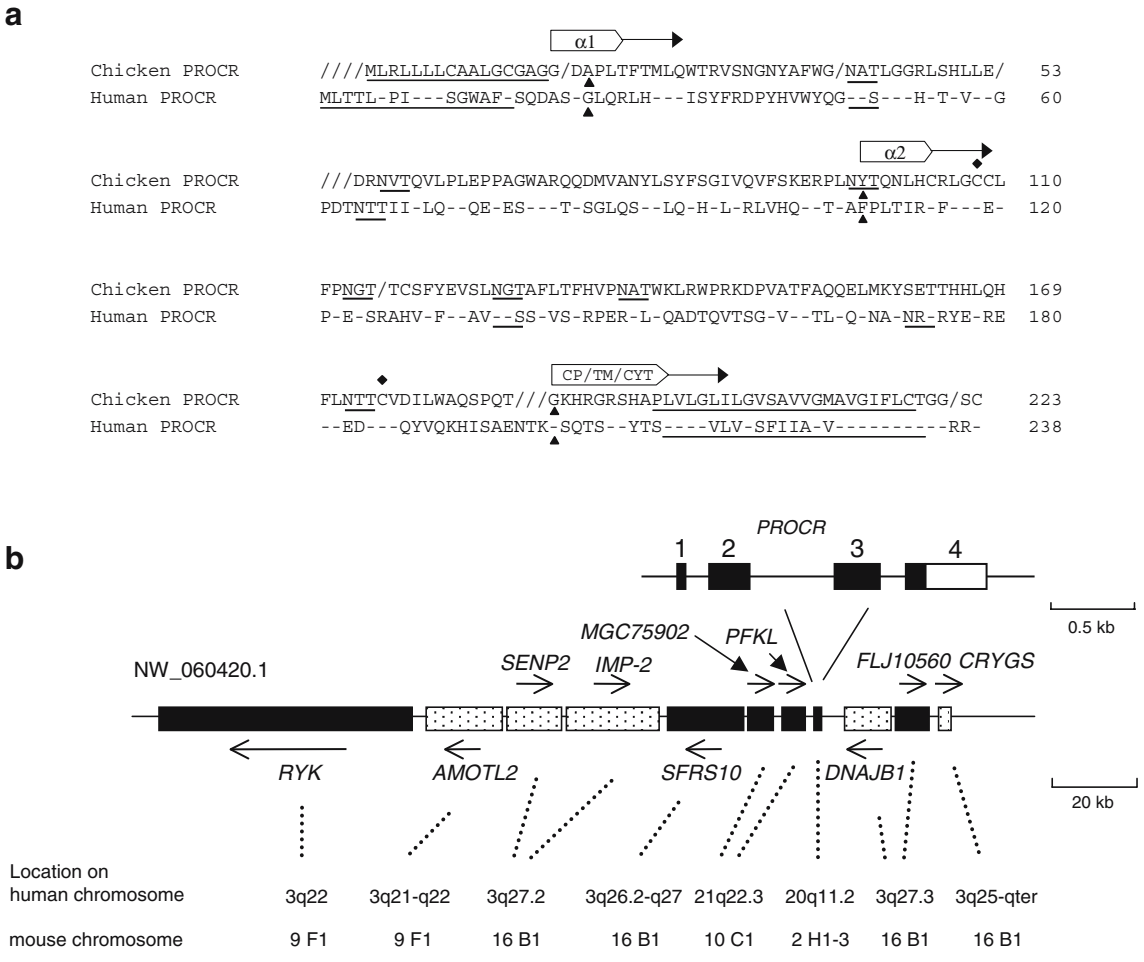

Fig. 5 **a** Comparison of the deduced amino acid sequences of chicken and human PROCR. The signs "–" and "/" indicate identity with the top sequence and absence of residues, respectively. Predicted signal peptides and TM are *underlined*. *Triangles* indicate exon–intron boundaries. *Diagonals* indicate conserved cysteine residues in the α2 domain. Potential *N*-linked glycosylation sites are indicated by *thick underlines*. The chicken PROCR cDNA sequence has been submitted to the DDBJ/EMBL/GenBank databases under accessions number AB204800. **b** Physical map surrounding *PROCR*. Exon–intron organization of chicken *PROCR* was deduced from the comparison of genomic and cDNA sequences. *Numbered* and *filled boxes* in *PROCR* indicate exons and coding regions, respectively. *Empty boxes* indicate 5′- and 3′-UTR. Other genes are indicated as *boxes with arrows* indicating transcriptional orientation. *Black boxes* indicate genes for which full-length cDNA sequences are available. All other genes are indicated as *stippled boxes*. Shown at the *bottom* of the figure are the chromosomal locations of the corresponding human and mouse genes

PROCR spans ~1.8 kb and has four exons (Fig. 5b). Exon 1 encodes the 5′-UTR, the signal peptide, and the first three residues of the α1 domain. Exons 2 and 3 encode the α1 and α2 domains, respectively, and exon 4 encodes the CP, TM, CYT, and 3′-UTR. All exon–intron boundaries conformed to the canonical GT/AG rule. Like human and mouse *PROCR* genes, RNA splicing took place between the first and second base of the junctional codons in all boundaries. We mapped the transcription initiations sites of chicken *PROCR* by sequencing 13 clones of 5′-RACE products. Twelve out of 13 clones started at the guanine residue located 18 bp upstream from the first adenine residue of the translation initiation site, indicating that this position constitutes a major transcription initiation site. A sequence that may function as a thrombin response element (minimum core sequence CCACCC) (Scarpati and DiCorleto 1996) was located at positions –234 to –229 (not shown). The 5′-flanking region of chicken *PROCR* contained three SP1-binding sites (–459 to –451, –388 to –380 and –200 to –192) but, unlike human but similar to mouse *PROCR*, possessed no canonical TATA or CCAAT box.

Discussion

In the present study, we showed that chickens have at least two families of MHC class-I-like genes: *CD1* and *PROCR*. Modern birds including chickens are thought to have originated from archosaurians derived from primitive diapsids. On the other hand, mammals are thought to have originated from primitive synapsids. Thus, the existence of *CD1* and *PROCR* in mammals and birds indicates that a common ancestor of diapsids and synapsids, hence that of all amniotes (Meyer and Zardoya 2003), had these two families of class-I-like genes. Interestingly, phylogenetic analysis indicates that *CD1* and *PROCR* are most closely related to each other among known families of class-I-like genes and that this relationship is supported by high bootstrap values (Fig. 2), consistent with the observation that the crystal structure of PROCR α1 and α2 domains (Oganesyan et al. 2002) resembles that of CD1d (Zeng et al. 1997). Thus, *CD1* and *PROCR* constitute a subfamily of class-I-like genes that predates the emergence of mammals. Can the origin of a common ancestor of *CD1* and *PROCR* be traced back to even lower classes of vertebrates? To address this issue, we performed computer searches using the genome databases of Fugu (JGI assembly v3.0), zebrafish (Ensembl 27.4b.1), and *Xenopus tropicalis* (JGI assembly v3.0). This analysis failed to provide any evidence for the existence of *CD1* or *PROCR* in these organisms (unpublished data). However, this may be accounted for by the incompleteness of the databases, sequence divergence that exceeds the limits of detection, and/or gene loss specific to animals subjected to genome sequencing. Thus, additional work will be required to draw conclusions as to the phylogenetic origin of the *CD1/PROCR* family.

The most interesting and unexpected observation made in this work is that chicken *CD1* genes are encoded in MHC *B* system adjacent to the class III region (Figs. 3 and 4). In mammals including humans, mice, and rats, *CD1* genes are located outside the MHC (Albertson et al. 1988; Matsuura et al. 1999; Moseley et al. 1989). The most straightforward interpretation of this observation is that the chicken MHC retains original gene configuration: *CD1* genes were originally located in the MHC and secondarily translocated to the outside of the MHC in the mammalian lineage. This implies that *CD1* genes emerged within the MHC by tandem duplication of class I genes. Whether an original class I gene was involved in the presentation of lipids like CD1 or of peptides like classical MHC class I is an open question. Because classical class I genes have been identified in all classes of jawed vertebrates (Flajnik and Kasahara 2001; Kelley et al. 2005), whereas *CD1* genes have been identified thus far only in mammals and chickens, we favor the view that peptide-binding class I molecules emerged earlier than CD1. In mammals, *CD1* genes are located in one of the four MHC paralogous regions presumed to have emerged by two successive rounds of block duplication that took place close to the origin of jawed vertebrates (Kasahara 1999). Because the α3 domain of CD1 is equally distant from those of MHC class I and class II molecules, it has been proposed that *CD1* genes emerged at the time when class I and class II genes diverged (Martin et al. 1986). On the basis of this observation, we proposed previously that *CD1* and MHC class I genes may have diverged by block duplication that formed the MHC paralogous regions (Kasahara et al. 1997). The localization of chicken *CD1* genes to the MHC, however, favors the idea that they diverged by tandem rather than block duplication. Nevertheless, we cannot rule out the possibility that *CD1* was secondarily translocated to the MHC in the chicken lineage. While the present paper was under review, two papers describing the linkage of chicken *CD1* to the MHC appeared (Miller et al. 2005; Salomonsen et al. 2005).

In mammals, CD1 molecules are classified into two major groups with distinct immune functions (Calabi et al. 1989b; Dascher and Brenner 2003; Park and Bendelac 2000). In man, CD1a, CD1b, CD1c, and possibly CD1e are group 1 molecules, and CD1d is a member of the group 2 CD1 molecules. Mice and rats have only group 2 CD1 molecules (Balk et al. 1991; Ichimiya et al. 1994), but rabbit and sheep have both groups of CD1 molecules (Calabi et al. 1989a; Dascher et al. 1999; Hayes and Knight 2001), indicating that group 1 *CD1* genes were lost in a rodent lineage (Dascher and Brenner 2003). Phylogenetic analysis indicates that chicken *CD1.1* and *CD1.2* are neither group 1- nor group 2-like and that the duplication event that gave rise to chicken *CD1.1* and *CD1.2* took place independently from the one that created the two groups of mammalian *CD1* genes (Fig. 2). In mammals, group 1 CD1 molecules interact with "mainstream" T cells, whereas group 2 CD1 molecules are recognized by natural killer T (NKT) cells (Vincent et al. 2003). Because the development of NKT cells is impaired in mice deficient in group 2 CD1 molecules (Chen et al. 1997; Mendiratta et al. 1997), chickens, which do not have an ortholog of

mammalian group 2 CD1, may lack authentic NKT cells. It is, however, also possible that functional specialization into two groups occurred in the chicken CD1 system by convergent evolution.

Another family of class-I-like genes identified in the chicken was *PROCR* (Fig. 5), which encodes the endothelial protein C receptor expressed on endothelial cells of large veins and arteries (Fukudome and Esmon 1994). Unlike most other members of the class I family, this receptor, which is involved in the protein C anticoagulant pathway, is a membrane glycoprotein with only two extracellular domains (α1 and α2). Like its mammalian counterpart, chicken *PROCR* has only two extracellular domains and its exon–intron structure is essentially identical to those of human and mouse counterparts (Hayashi et al. 1999; Liang et al. 1999; Simmonds and Lane 1999). In humans and mice, *PROCR* has been localized to 20q11.2 and 2H1-3, respectively (Hayashi et al. 1999; Mincheva et al. 1995; Simmonds and Lane 1999), whereas the draft genome assembly predicts that chicken *PROCR* is on chromosome 9. The region of the chicken genome corresponding to human 20q11 and mouse 2H1-3 is on chromosome 20. Conversely, none of the genes in the vicinity of chicken *PROCR* has its counterpart on human 20q11.2 or mouse chromosome 2 (Fig. 5b). Thus, if the chicken draft genome assembly is correct, mammalian and chicken *PROCR* are located in nonorthologous chromosomal segments. If the translocation of *PROCR* from the MHC took place only once in a common ancestor of mammals and chickens, this gene must have undergone another round of translocation in either a mammalian or avian lineage. Alternatively, translocation of *PROCR* may have occurred independently in mammalian and avian lineages, thus resulting in localization in nonorthologous chromosomal segments.

Despite extensive database searches, we were unable to identify *MICA/B*, *Mill*, *FCGRT*, *AZGP1*, *MR1*, *HFE*, or *RAET/ULBP* families of class-I-like genes in the chicken draft genome. This is likely because some of them emerged first in the mammalian lineage. For example, *FCGRT*, which encodes the heavy chain of the neonatal IgG Fc receptor involved in the transfer of maternal IgG across intestinal epithelia and yolk sac membranes (Simister and Mostov 1989), seems to be a mammalian invention because the transfer of maternal IgY, the chicken counterpart of mammalian IgG, to embryos across yolk sac membranes is mediated by a receptor structurally unrelated to the neonatal IgG Fc receptor (West et al. 2004). However, others may have simply eluded detection because the available draft genome sequence was incomplete, their sequences were too divergent, or they were lost specifically in the chicken as a result of domestication or for other reasons.

Acknowledgements We thank Dr. Nagata, Ms. Kuno, and Ms. Muraguchi for technical assistance. This work was supported by Grants-in-Aid for Scientific Research from The Ministry of Education, Culture, Sports, Science and Technology of Japan; Uehara Memorial Foundation; and The Naito Foundation.

References

Albertson DG, Fishpool R, Sherrington P, Nacheva E, Milstein C (1988) Sensitive and high resolution in situ hybridization to human chromosomes using biotin labelled probes: assignment of the human thymocyte CD1 antigen genes to chromosome 1. EMBO J 7:2801–2805

Araki T, Gejyo F, Takagaki K, Haupt H, Schwick HG, Buergi W, Marti T, Schaller J, Rickli E, Brossmer R, Atkinson PH, Putnam FW, Schmid K (1988) Complete amino acid sequence of human plasma Zn-α2-glycoprotein and its homology to histocompatibility antigens. Proc Natl Acad Sci U S A 85:679–683

Bahram S (2000) *MIC* genes: from genetics to biology. Adv Immunol 76:1–60

Bahram S, Bresnahan M, Geraghty DE, Spies T (1994) A second lineage of mammalian major histocompatibility complex class I genes. Proc Natl Acad Sci U S A 91:6259–6263

Balk SP, Bleicher PA, Terhorst C (1991) Isolation and expression of cDNA encoding the murine homologues of CD1. J Immunol 146:768–774

Bloom SE, Bacon LD (1985) Linkage of the major histocompatibility (B) complex and the nucleolar organizer in the chicken. Assignment to a microchromosome. J Heredity 76:146–154

Calabi F, Milstein C (1986) A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. Nature 323:540–543

Calabi F, Belt KT, Yu CY, Bradbury A, Mandy WJ, Milstein C (1989a) The rabbit *CD1* and the evolutionary conservation of the *CD1* gene family. Immunogenetics 30:370–377

Calabi F, Jarvis JM, Martin L, Milstein C (1989b) Two classes of *CD1* genes. Eur J Immunol 19:285–292

Chen YH, Chiu NM, Mandal M, Wang N, Wang CR (1997) Impaired NK1+ T cell development and early IL-4 production in CD1-deficient mice. Immunity 6:459–467

Dascher CC, Brenner MB (2003) Evolutionary constraints on *CD1* structure: insights from comparative genomic analysis. Trends Immunol 24:412–418

Dascher CC, Hiromatsu K, Naylor JW, Brauer PP, Brown KA, Storey JR, Behar SM, Kawasaki ES, Porcelli SA, Brenner MB, LeClair KP (1999) Conservation of a *CD1* multigene family in the guinea pig. J Immunol 163:5478–5488

Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo RJ, Ellis MC, Fullan A, Hinton LM, Jones NL, Kimmel BE, Kronmal GS, Lauer P, Lee VK, Loeb DB, Mapa FA, McClelland E, Meyer NC, Mintier GA, Moeller N, Moore T, Morikang E, Prass CE, Quintana L, Starnes SM, Schatzman RC, Brunke KJ, Drayna DT, Risch NJ, Bacon BR, Wolff RK (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13:399–408

Fillon V, Morisson M, Zoorob R, Auffray C, Douaire M, Gellin J, Vignal A (1998) Identification of 16 chicken microchromosomes by molecular markers using two-colour fluorescence in situ hybridization (FISH). Chromosome Res 6:307–313

Flajnik MF, Kasahara M (2001) Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity 15:351–362

Frangoulis B, Park I, Guillemot F, Severac V, Auffray C, Zoorob R (1999) Identification of the *Tapasin* gene in the chicken major histocompatibility complex. Immunogenetics 49:328–337

Fukudome K, Esmon CT (1994) Identification, cloning, and regulation of a novel endothelial cell protein C/activated protein C receptor. J Biol Chem 269:26486–26491

Hagios C, Koch M, Spring J, Chiquet M, Chiquet-Ehrismann R (1996) Tenascin-Y: a protein of novel domain structure is secreted by differentiated fibroblasts of muscle connective tissue. J Cell Biol 134:1499–1512

Hashimoto K, Hirai M, Kurosawa Y (1995) A gene outside the human MHC related to classical HLA class I genes. Science 269:693–695

Hayashi T, Nakamura H, Okada A, Takebayashi S, Wakita T, Yuasa H, Okumura K, Suzuki K (1999) Organization and chromosomal localization of the human endothelial protein C receptor gene. Gene 238:367–373

Hayes SM, Knight KL (2001) Group 1 *CD1* genes in rabbit. J Immunol 166:403–410

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang SP, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobbson L, Kerje S, Andersson L, Crooijmans RP, Aerts J, van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MD, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WR, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GK, Wang J, Liu B, Yu J, Yang H, Nefedov M, Koriabine M, Dejong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, von Mering C, Zdobnov EM et al (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716

Ichimiya S, Kikuchi K, Matsuura A (1994) Structural analysis of the rat homologue of *CD1*. Evidence for evolutionary conservation of the *CD1D* class and widespread transcription by rat cells. J Immunol 153:1112–1123

Kasahara M (1999) The chromosomal duplication model of the major histocompatibility complex. Immunol Rev 167:17–32

Kasahara M, Nakaya J, Satta Y, Takahata N (1997) Chromosomal duplication and the emergence of the adaptive immune system. Trends Genet 13:90–92

Kasahara M, Watanabe Y, Sumasu M, Nagata T (2002) A family of MHC class I-like genes located in the vicinity of the mouse leukocyte receptor complex. Proc Natl Acad Sci U S A 99:13687–13692

Kaufman J, Bumstead N, Miller M, Riegert P, Salomonsen J (1995) The chicken class II α gene is located outside the *B* complex. In: Davison TF, Bumstead N, Kaiser P (eds) Advances in avian immunology research. Carfax, Abingdon, pp 119–127

Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck D, Salomonsen J (1999a) Gene organisation determines evolution of function in the chicken MHC. Immunol Rev 167:101–117

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999b) The chicken *B* locus is a minimal essential major histocompatibility complex. Nature 401:923–925

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. Immunogenetics 56:683–695

Klein J (1986) Natural history of the major histocompatibility complex. Wiley, New York

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Liang Z, Rosen ED, Castellino FJ (1999) Nucleotide structure and characterization of the murine gene encoding the endothelial cell protein C receptor. Thromb Haemost 81:585–588

Martin LH, Calabi F, Milstein C (1986) Isolation of *CD1* genes: a family of major histocompatibility complex-related differentiation antigens. Proc Natl Acad Sci U S A 83:9154–9158

Maruyama K, Sugano S (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene 138:171–174

Matsuura A, Kinebuchi M, Katabami S, Chen HZ, Yamada K, Yoshida MC, Horie M, Watanabe TK, Kikuchi K, Sato N (1999) Correction and confirmation of the assignment of *Cd1d* the evolutionarily conserved *CD1D* class gene to rat chromosome 2q34 and its relationship to human and mouse loci. Cytogenet Cell Genet 86:323–324

Mendiratta SK, Martin WD, Hong S, Boesteanu A, Joyce S, Van Kaer L (1997) CD1d1 mutant mice are deficient in natural T cells that promptly produce IL-4. Immunity 6:469–477

Meyer A, Zardoya R (2003) Recent advances in the (molecular) phylogeny of vertebrates. Annu Rev Ecol Evol Syst 34:311–338

Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J, Zoorob R, Briles WE (2004) 2004 Nomenclature for the chicken major histocompatibility (*B* and *Y*) complex. Immunogenetics 56:261–279

Miller MM, Wang C, Parisini E, Coletta RD, Goto RM, Lee SY, Barral DC, Townes M, Roura-Mir C, Ford HL, Brenner MB, Dascher CC (2005) Characterization of two avian MHC-like genes reveals an ancient origin of the CD1 family. Proc Natl Acad Sci U S A 102:8674–8679

Mincheva A, Rothbarth K, Werner D, Lichter P (1995) Assignment of the gene encoding centrosome-associated protein CCD41 to mouse chromosome 2H. Mamm Genome 6:444

Moseley WS, Watson ML, Kingsmore SF, Seldin MF (1989) *CD1* defines conserved linkage group border between human chromosome 1 and mouse chromosomes 1 and 3. Immunogenetics 30:378–382

Oganesyan V, Oganesyan N, Terzyan S, Qu D, Dauter Z, Esmon NL, Esmon CT (2002) The crystal structure of the endothelial protein C receptor and a bound phospholipid. J Biol Chem 277:24851–24854

Park SH, Bendelac A (2000) CD1-restricted T-cell responses and microbial infection. Nature 406:788–792

Radosavljevic M, Bahram S (2003) In vivo immunogenetics: from *MIC* to *RAET1* loci. Immunogenetics 55:1–9

Salomonsen J, Marston D, Avila D, Bumstead N, Johansson B, Juul-Madsen H, Olesen GD, Riegert P, Skjodt K, Vainio O, Wiles MV, Kaufman J (2003) The properties of the single chicken MHC classical class II alpha chain (*B-LA*) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. Immunogenetics 55:605–614

Salomonsen J, Sorensen MR, Marston DA, Rogers SL, Collen T, van Hateren A, Smith AL, Beal RK, Skjodt K, Kaufman J (2005) Two *CD1* genes map to the chicken MHC, indicating that *CD1* genes are ancient and likely to have been present in the primordial MHC. Proc Natl Acad Sci U S A 102:8668–8673

Sandoval IV, Bakke O (1994) Targeting of membrane proteins to endosomes and lysosomes. Trends Cell Biol 4:292–297

Scarpati EM, DiCorleto PE (1996) Identification of a thrombin response element in the human platelet-derived growth factor B-chain (c-sis) promoter. J Biol Chem 271:3025–3032

Simister NE, Mostov KE (1989) An Fc receptor structurally related to MHC class I antigens. Nature 337:184–187

Simister NE, Rees AR (1985) Isolation and characterization of an Fc receptor from neonatal rat small intestine. Eur J Immunol 15:733–738

Simmonds RE, Lane DA (1999) Structural and functional implications of the intron/exon organization of the human endothelial cell protein C/activated protein C receptor (EPCR) gene: comparison with the structure of CD1/major histocompatibility complex alpha1 and alpha2 domains. Blood 94:632–641

Sugita M, Cernadas M, Brenner MB (2004) New insights into pathways for CD1-mediated antigen presentation. Curr Opin Immunol 16:90–95

Tanabe H, Kupper K, Ishida T, Neusser M, Mizusawa H (2005) Inter- and intra-specific gene-density-correlated radial chromosome territory arrangements are conserved in Old World monkeys. Cytogenet Genome Res 108:255–261

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Todorov PT, McDevitt TM, Meyer DJ, Ueyama H, Ohkubo I, Tisdale MJ (1998) Purification and characterization of a tumor lipid-mobilizing factor. Cancer Res 58:2353–2358

Treiner E, Duban L, Bahram S, Radosavljevic M, Wanner V, Tilloy F, Affaticati P, Gilfillan S, Lantz O (2003) Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. Nature 422:164–169

Vincent MS, Gumperz JE, Brenner MB (2003) Understanding the function of CD1-restricted T cells. Nat Immunol 4:517–523

Watanabe Y, Maruoka T, Walter L, Kasahara M (2004) Comparative genomics of the *Mill* family: a rapidly evolving MHC class I gene family. Eur J Immunol 34:1597–1607

West AP Jr, Herr AB, Bjorkman PJ (2004) The chicken yolk sac IgY receptor, a functional equivalent of the mammalian MHC-related Fc receptor, is a phospholipase A2 receptor homolog. Immunity 20:601–610

Zeng Z, Castano AR, Segelke BW, Stura EA, Peterson PA, Wilson IA (1997) Crystal structure of mouse CD1: an MHC-like fold with a large hydrophobic binding groove. Science 277:339–345

Zou Z, Nomura M, Takihara Y, Yasunaga T, Shimada K (1996) Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: a novel cDNA family encodes cell surface proteins sharing partial homology with MHC class I molecules. J Biochem (Tokyo) 119:319–328